

## Association for Information Systems AIS Electronic Library (AISeL)

---

PACIS 2007 Proceedings

Pacific Asia Conference on Information Systems  
(PACIS)

---

2007

# Visualized Cognitive Knowledge Map Integration for P2P Networks

Fu-ren Lin

*National Tsing Hua University, frlin@mx.nthu.edu.tw*

Jen-Hung Yu

*National Tsing Hua University, jhyu@gmail.com*

Follow this and additional works at: <http://aisel.aisnet.org/pacis2007>

---

### Recommended Citation

Lin, Fu-ren and Yu, Jen-Hung, "Visualized Cognitive Knowledge Map Integration for P2P Networks" (2007). *PACIS 2007 Proceedings*. 69.

<http://aisel.aisnet.org/pacis2007/69>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

## 91. Visualized Cognitive Knowledge Map Integration for P2P Networks

Fu-ren Lin  
National Tsing Hua University  
[frlin@mx.nthu.edu.tw](mailto:frlin@mx.nthu.edu.tw)

Jen-Hung Yu  
National Tsing Hua University  
[jhyu@gmail.com](mailto:jhyu@gmail.com)

### Abstract

*In the knowledge management field, knowledge map created under the client-server architecture has been widely used to direct the knowledge sharing process. Peer-to-peer (P2P) architecture has been practicable for file sharing, distributed computing, instant messaging, etc. by virtue of the increases of Internet bandwidth and personal computer capability. P2P architecture attracts researchers and practitioners to study knowledge sharing issues in its autonomy and self-organization characters. It calls for academic efforts to design knowledge management supporting system as that for client-server architecture may not be applicable. This study proposes a visualized cognitive knowledge map integration system to facilitate knowledge management on P2P networks. By using the SOM (Self-Organized Map)-like model, called Egocentric SOM (ESOM), the prototyping system can merge the external knowledge under a focal peer's knowledge structure and present the cognitive knowledge map visually. In evaluating the proposed integration method, this study allocates abstracts of industrial research reports from an industrial technology research institute according to their corresponding author peers and generates individual cognitive knowledge maps. The results from the evaluation experiments reveal that ESOM is capable to retain individual peer's knowledge structure while articulating with that of other peers in the cognitive knowledge network.*

**Keywords:** Self-Organizing Map (SOM), Knowledge Map, Peer-to-Peer (P2P) Architecture, Egocentric SOM (ESOM)

### Introduction

Knowledge management (KM) has been widely recognized as an important activity for retaining and increasing organizational competitiveness. Client-server architecture operated on centralized and homogeneous knowledge databases or repositories has been widely used for KM systems both in research and practice. For better managing knowledge, a knowledge map created under client-server architecture also has been widely used to provide a unified structure to represent knowledge embedded in an organization and to assist its members to acquire and disseminate knowledge.

With increasing Internet bandwidth and the personal computer capability, the peer-to-peer (P2P) architecture is getting practicable for file sharing, distributed computing, instant messaging and so on. Against the client-server architecture, the P2P architecture emphasizes self-organization by peers instead of administration by a central server, and provides more elastic knowledge structure decided by peers instead of an inflexible unified knowledge structure limited by a central server. By contrast, the P2P architecture is more similar to knowledge acquisition and dissemination process that knowledge workers autonomously decide the way to handle their knowledge, and may be more suitable than the client-server architecture for KM. Therefore, it attracts more attention from researchers in KM discipline (Castano et al. 2003; Mangisengi and Essmayr 2003).

Knowledge map methodologies proposed based on the client-server architecture may not be

suitable for the P2P architecture. Firstly, there is no central server which can take charge of gathering all information to create knowledge map in the P2P architecture. Secondly, the distribution of knowledge in the P2P architecture is more skewed than in the client-server architecture due to the widely spread of P2P networks. Thus, this paper proposes the architecture of visualized cognitive knowledge map integration system for P2P networks. It is hard to guarantee that all peers directly interconnect with each other and keep online all the time in P2P networks. Therefore, we use the cognitive knowledge map rather than a global knowledge map to denote the knowledge network that a peer recognizes in this study. Different from tree-based or category-based knowledge maps, a cognitive knowledge map created by the SOM (Self-Organizing Map)-like algorithm, called Egocentric SOM, proposed in this study is a visualized two-dimensional rectangle map which preserves individual knowledge structure and presents inter relationship among knowledge artifacts, *e.g.*, documents.

The first characteristic of the proposed visualized cognitive knowledge map integration system is that it combines the individuals' knowledge map without a pre-decided common term vector. The second is that by using an individual knowledge map as an initiative map via Egocentric SOM, it not only retains individual peers' original knowledge structures but also represents the correct distribution of individual knowledge as much as possible. The third is that based on a two-dimensional rectangle map, it provides the visualization ability to assist peer users to understand and retrieve the knowledge distribution of a P2P network. Therefore, this system can provide individual peers the visualized cognitive knowledge map which conforms to the nature of knowledge sharing, and assist them to retrieve knowledge from other peers within or across organizational boundaries in a visualized way. We demonstrate the visualized cognitive knowledge map integration system using abstracts of industrial research reports collected from the Industrial Economics and Knowledge (IEK) Center at the Industrial Technology Research Institute (ITRI), Taiwan.

Section 2 of this study reviews the related SOM-like algorithm and P2P knowledge management literatures. Section 3 proposes the framework of visualized cognitive knowledge map integration system. The ESOM clustering algorithm is described in Section 4. In section 5, we evaluate the performance of this system. Section 6 concludes this study.

## **Related Works**

This study proposes a knowledge map integration system suitable for clustering knowledge on P2P networks to generate egocentric knowledge structure. It is based on mechanisms from self-organizing map related techniques described in Subsection 2.1. Recent efforts on distributed knowledge management especially on P2P networks are also reviewed in Subsection 2.2.

### ***Self-Organizing Map (SOM)***

Self-Organizing Map (SOM) proposed by Kohonen (1982) is an unsupervised artificial neural network model which can be trained by feeding input data repeatedly without any extra guidance. The major advantage of SOM is that it can map high-dimensional input data into a two-dimensional grid topology which provides the visualization ability to represent the relationship among input data; hence, SOM has been introduced to solving document clustering problems.

### ***SOM Algorithm***

The original data, such as documents, are transformed to vector space model (VSM) (Salton,

Wong, and Yang 1975) as inputs to the SOM model. A VSM is a  $m \times n$  matrix. The number of columns,  $n$ , denotes the total number of distinct terms in a document collection, and the number of rows,  $m$ , denotes the total number of documents, where each document is represented as a vector of terms. The weight of each distinct term in each document,  $w_{ij}$ , is simply measured by binary numbers. Namely,  $w_{ij} = 0$  denotes a distinct term  $j$  does not appear in document  $i$ , whereas  $w_{ij} = 1$  denotes it does.

*tfidf* weight,  $w_{ij} = tf_{ij} \cdot \log \frac{N}{n}$  is generally used to determine a term's importance in distinguishing documents (Luhn 1957; Sparck Jones 1972), where  $w_{ij}$  is the weight of distinct term  $t_j$  in a document  $D_i$ ,  $tf_{ij}$  is the frequency of  $t_j$  appears in document  $D_i$ ,  $N$  is the number of documents in the collection, and  $n$  is number of documents where term  $t_j$  appears at least once. In a word, the higher weight of a term the higher discriminating power the term has. After calculating the weight for each distinct term, we select the top  $N$  terms to form the VSM.  $N$  is often determined by trial and error on the basis of balancing the computing cost and discriminating power.

A SOM algorithm consists of initialization and training phases. In the initialization phase, the number of the units and the topology of output layer are determined first, and then the reference vector of each unit  $m_i$  of output layer is initiated, where each reference vector  $m_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{in}]^T \in \mathbb{R}^n$ .

In the training phase, first, a term vector,  $x$ , is randomly selected from document corpus to finds the best-matching unit (BMU) from all units of output layer for  $x$ . The best-matching unit is identified as  $\|x - m_c\| = \min_i \|x - m_i\|$ , where  $c$  is the BMU,  $i$  is an units of output layer and  $\|x - m_i\|$  is the Euclidean distance between  $x$  and  $i$ . Second, it activates the BMU and other units which are topographically close to the BMU through the formula  $m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]$ , where  $t = 0, 1, 2, \dots$  is an integer to denote the discrete-time coordinate,  $h_{ci}(t)$  is called neighborhood function shown as  $h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right)$ , where  $\alpha(t)$  is the learning rate ( $0 < \alpha(t) < 1$ ) decaying with time,  $r_c \in \mathbb{R}^2$  and  $r_i \in \mathbb{R}^2$  are the location vectors of nodes  $c$  and  $i$ , and  $\sigma(t)$  defines the width of the kernel. In the neighborhood function,  $\|r_c - r_i\|^2$  is used to measure the distance from the units to the BMU; obviously,  $h_{ci}(t)$  is inverse proportion with distance of these two units.  $\alpha(t)$  and  $\sigma(t)$  are monotonically decreasing functions of time; therefore, the training phase will be convergent as time goes by. In practice, it usually selects a specific number of training iterations as the stop criterion for stopping training.

By the aforementioned SOM algorithm, the input data with similar concepts will be projected into the same zone, so that SOM can facilitate users to observe the relationship among input data. However, there are three major limitations which make SOM unsuitable for real world applications. First, the topology and the number of units of output layer should be determined prior to the training phase and is fixed throughout the process. Second, SOM is a single layer map which cannot be used to represent the hierarchical relationship between

input data. Third, the decayed learning rate may reduce to a very small value which prohibits SOM from further improving.

Many researches have addressed these aforementioned SOM defects, and proposed different SOM-like models to overcome these weak points. Three major enhanced SOM-like models have been proposed, which are Growing Hierarchical Self-Organizing Map (Rauber, Merkl, and Dittenbach 2002), Growing Neural Gas (Fritzke 1995), and Dynamic Adaptive Self-organizing Hybrid Model (HUNG and Wermter 2003).

### *Growing Hierarchical Self-Organizing Map (GHSOM)*

The Growing Hierarchical Self-Organizing Map (GHSOM) (Rauber, Merkl, and Dittenbach 2002) is composed of individual growing self-organizing maps to form a hierarchical artificial neural network model. By providing unit-growing function in training phase, GHSOM can adjust the topology and the number of units of a map according to input data and parameters automatically. Moreover, GHSOM can expand units which are too diversely by a new map in the sub-layer, and therefore it is capable of tracing diverse units in the hierarchical way.

GHSOM starts with a virtual map,  $M_0$ , which has only one virtual unit called  $U_0$  in Layer 0.  $U_0$  represents the total input data which will be used to train the model later. For each map in each layer except  $M_0$ , GHSOM initiates it with a  $2 \times 2$  grid and trains it by SOM algorithm mentioned in Subsection 2.1.1.

The unit-growing function is applied to each training epoch. At each training epoch, firstly, it identifies the unit with the most quantization error (QE) called unit  $E$ ; secondly, it selects the most irrelevant neighbor against unit  $E$  called unit  $D$ ; finally, it inserts a row or a column of units between units  $E$  and  $D$ . Afterward, parameters of the map will be reset by their initial value and the next round of training epoch will proceed. This process will stop if the mean quantization error (MQE) of the current map is less than the fraction  $\tau_1$  of the QE of its parent unit. After the process of training each map, GHSOM will find out the units that their QE are more than the fraction  $\tau_2$  of QE of unit  $U_0$  and expand these units by maps in the sub-layer.

### ***Peer-to-Peer Knowledge Management (P2PKM)***

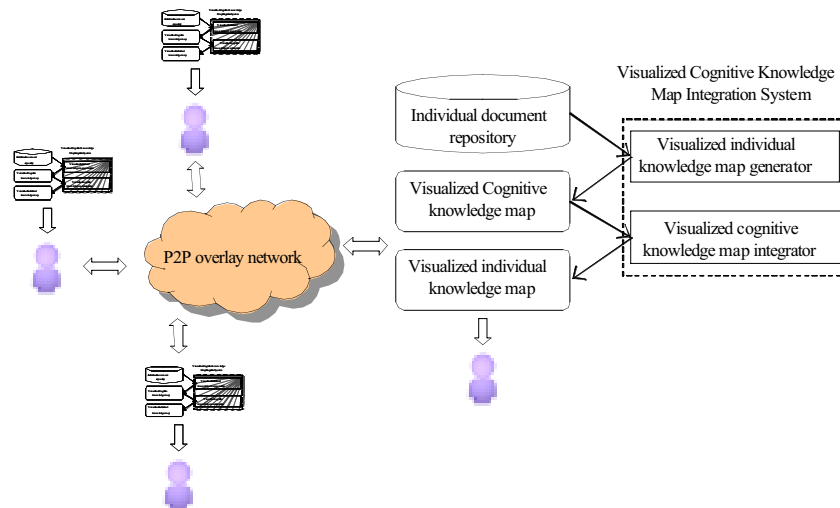
Knowledge management (KM) consisting of creating, codifying and disseminating knowledge within the organization has been widely recognized as an important activity for retaining or increasing organizational competitiveness. Current KM research mostly focus on investigating knowledge activities on client-server architecture which uses centralized homogeneous knowledge databases or repositories to provide unified knowledge structure for an organization. Such unified knowledge structure, often called knowledge map, ontology, categorization, or classification system, is meant to represent a shared conceptualization of corporate knowledge, and to enable communication and knowledge exchange across the entire organization (Bonifacio et al. 2002). We denote this unified knowledge structure as knowledge map in this study.

The enhancement of Internet bandwidth and personal computer capability enables peer-to-peer (P2P) architecture to serve as a knowledge sharing platform. Contrast to the client-server architecture, the P2P architecture emphasizes self-organization by peers rather than the administration by a central server, and provides more elastic knowledge structure decided by peers rather than an inflexible unified knowledge structure limited by the central server.

Thus, the P2P architecture is more suitable for knowledge sharing across professional communities where peers have high autonomy in the knowledge acquisition and dissemination process. The research in knowledge sharing on P2P networks is still in its infancy as shown in Table 1. This study focuses on a conceptual P2P architecture, and the proposed cognitive knowledge map integration is suitable for different P2P implementation protocols. The research outcome from this study on knowledge map integration will contribute to the attempt to generate cognitive knowledge map on P2P networks.

**Table 1: Summary of Related Works of Knowledge Management on P2P Networks**

References	Major works
Bonifacio et al. 2002	It proposes a P2P architecture called KEx according to the principles of distributed knowledge management (DKM). Each peer called KN can organize its knowledge by its perspective and the results are defined as context. Similar to ontology, context is used to describe the knowledge of peer by a hierarchical way and to do knowledge searching accomplished by the collaboration of peers.. KEx is implemented on the JXTA.
Castano et al. 2003	It proposes an architecture called HELIOS. By HELIOS toolkits, it provides the ability for peers to create individual ontology by themselves, and provides concept matching rather than traditional keyword search. The main characteristic of HELIOS is that peers not only control their own ontology, but also they can extend their own ontology by concept matching.
Mangisengi and Essmayr 2003	It provides an ontology-based P2P KM architecture. Every peer creates its local ontology in XML format, and bases on individual ontologies to conduct search. It also provides the ability to integrate different ontologies.



**Figure 1: Architecture of *VisCog***

## Architecture of Visualized Cognitive Knowledge Map Integration System (VisCog)

The architecture of the visualized cognitive knowledge map integration system called *VisCog* proposed in this study is shown in Figure 1. It embedded in a host peer takes charge of integrating neighbor peers' knowledge structures gathered through the P2P overlay network

with its own knowledge map. With the cognitive knowledge map created based on a host knowledge map, *VisCog* enables its host peer the ability to navigate the knowledge distribution on a P2P network based on the host peer's own knowledge structure. *VisCog* consists of the following components:

- (1) A *visualized individual knowledge map* is a single-layer GHSOM which represents the document clusters of an individual document repository and is used to facilitate the navigation of the knowledge distribution of an individual document repository. Also, it acts as an initiative map for the visualized cognitive knowledge map integrator.
- (2) A *visualized individual knowledge map generator* is responsible for transforming unstructured documents in an individual document repository into structured data and clustering these structured data using the GHOSM algorithm mentioned in subsection 2.1.2.
- (3) A *visualized cognitive knowledge map* is the output of the visualized cognitive knowledge integrator. It not only retains original individual knowledge structures, but also represents the distribution of knowledge among peers according to its cognitive knowledge network. Moreover, by using SOM-like algorithm, it provides the visualized ability for facilitating knowledge seekers to retrieve other peers' knowledge.
- (4) The *visualized cognitive knowledge map integrator* is the kernel of *VisCog* to integrate an individual peer's knowledge map with its neighbor peers'.

### ***Visualized Individual Knowledge Map Generator***

The visualized individual knowledge map generator is responsible for extracting the knowledge structure of a peer's document repository. It uses information retrieval techniques to transform unstructured documents into structured data, and applies a clustering algorithm to identify the knowledge structure hidden in a peer's document repository. It is split up into three specific functional components and described as follows.

- (1) *Keyword extraction*: Due to no explicit word boundary in Chinese sentences, the word segmentation is the first step in all Chinese information retrieval. Because the Chinese segment system developed by CKIP projects in the Academia Sinica (<http://ckip.iis.sinica.edu.tw/CKIP/>) can identify known terms by the dictionary-based method, and unknown terms by the linguistic and statistical approaches, we directly use CKIP to segment sentences into basic meaningful lexicons. Moreover, we prune lexicons which are not noun or are single-character. Because most of these lexicons cannot represent complete concepts of documents.

- (2) *Term Weighting*: After the keyword extraction stage, candidate keywords have been extracted but the amount of these is still too much. Therefore, we further use *tfidf* weight to prune terms which have less discriminating power. Modifying from basic *tfidf* weight, this study consider the effect of different document lengths, and uses the normalized *tfidf*

weight  $w_{ij} = \frac{tf_{ij}}{\sum tf_i} \cdot \log \frac{N}{n}$ , where  $w_{ij}$  is weight of distinct term  $t_j$  in document  $D_i$ ,  $tf_{ij}$  is

the frequency of term  $t_j$  appears in document  $D_i$ ,  $\sum tf_i$  is the amount of terms in document  $D_i$ ,  $N$  is the amount of documents in the collection, and  $n$  is the number of documents where term  $t_j$  appears at least once. After calculating *tfidf* weight for every candidate keyword in keyword repository, we sort these candidate keywords according to their *tfidf* weight. Finally, we use the first 40% of these keywords and the traditional VSM mentioned in subsection 2.1.1 to transform the original document repository into a document-term matrix.

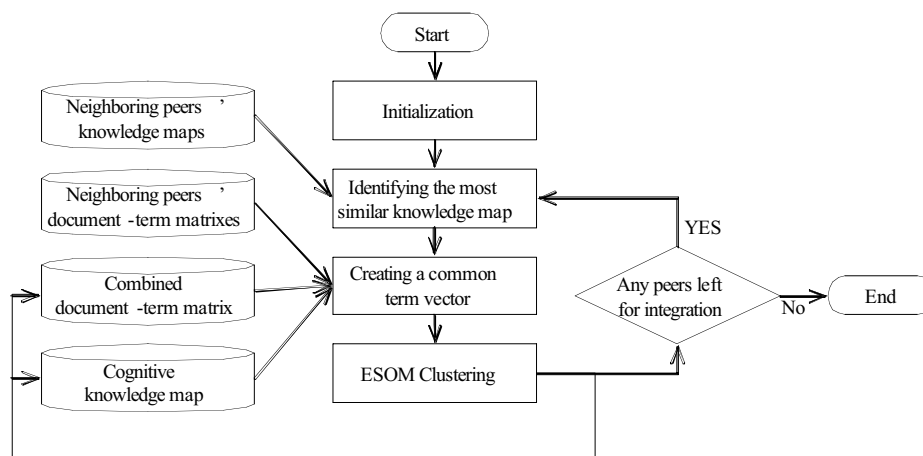
- (3) *Clustering*: After aforementioned pre-processing stages, documents represented as

document-term matrix in individual document repositories is suitable for clustering. Because GHSOM clustering algorithm provides the visualized capability by forming the result as a rectangle topology map and automatically adjusts the map size by the unit-growing function, we select GHSOM clustering algorithm for clustering documents. Moreover, due to the cognitive knowledge map is a single-layer map in this study, we set the expanded parameter  $\tau_2$  of GHSOM to 1 which means GHSOM will only generate a single-layer knowledge map.

### ***Visualized Cognitive Knowledge Map Integrator***

In the context of this study, the aim of the visualized individual knowledge map integrator is to construct a cognitive knowledge map which represents the cognition of knowledge distribution on a P2P network based on a focal peer's knowledge map. For the purpose, this study proposes the Egocentric SOM (ESOM) algorithm which emphasizes the preservation of the focal peer's knowledge structure while merging other peers' knowledge structure.

Like other SOM-like algorithms, ESOM also needs a predefined common term vector for a focal peer and its merged neighboring peer, yet knowledge workers are autonomous, they may use varied terms to create different term vectors for representing their own knowledge concepts. To solve the problem of different term vectors used by different peers and reduce the complexity of the integration, this study proposes a procedure of the visualized cognitive knowledge map integrator as shown in Figure 2 and describes it in the following subsections.



**Figure 2: The procedure of Visualized Cognitive Knowledge Map Integration**

### ***Initialization***

In the P2P environment, each peer only uses its own documents to create its individual knowledge map. These limited partial documents may result in biased document clusters in individual knowledge maps. To generate a common knowledge map across cognitive knowledge networks, it takes a great communication effort to exchange documents from corresponding peers' repositories. By contrast, the exchange of document-term matrix has at least two advantages. First, the document-term matrix only consists of a vector of terms and a matrix of numerals. Thus, it's efficient to transmit on the Internet. Second, the document-term matrix retains the elasticity for re-arranging document clusters. Hence, this study uses the document-term matrix rather than the individual knowledge map as the exchange object among peers, and initializes the combined document-term matrix with the focal peer's document-term matrix. Because ESOM uses the initial knowledge map as the framework to proceed with the integration, this study uses the focal peer individual knowledge map to



initialize the cognitive knowledge map.

#### *Identifying the Most Similar Knowledge Map*

With the nature of the P2P architecture, peers are autonomous and therefore terms used for representing their own knowledge may vary. This situation increases the complexity of the integration. To reduce the complexity of the integration, we merge only one peer at each round and the integration sequence follows the knowledge structure similarity between the focal peer and neighboring peers separately. That is, the focal peer merges a neighboring peer which has the most similar knowledge structure first.

Because we use knowledge map to represent peers' knowledge structure in this study, we measure the knowledge structure similarity between the focal and neighboring peers by comparing their knowledge maps, respectively. Due to their map sizes may differ, we identify a representative unit for every map before proceeding to compare. The representative unit is the average of all units of the map. The equation used to calculate the map similarity is

denoted as  $M_{ij} = (n(i \cap j) / n(i)) \cdot \left( n(i \cap j) / \sqrt{\sum_k (w_{ik} - w_{jk})^2} \right)$ , where  $n(i)$  is the number of

terms used in the cognitive knowledge map,  $n(i \cap j)$  is the number of the joint terms from the cognitive and the neighboring peer's knowledge maps,  $w_{ik}$  and  $w_{jk}$  are the weight of the identical term in the cognitive and the neighboring peer's knowledge maps, respectively, and

$\sqrt{\sum_k (w_{ik} - w_{jk})^2}$  is the Euclidean distance between two knowledge maps. This equation

emphasizes that a neighboring peer which has more identical terms and closer term weights with the focal peer will get higher map similarity.

#### *Creating a Common Term Vector*

Since term vectors used by the focal and merged neighboring peers may differ, ESOM cannot proceed to the integration process without aligning both vectors. To tackle this issue, *VisCog* generates a common term vector by the union of term vectors of both peers in advance. Then, the resulting common term vector is used by the combined and the neighboring peer's document-term matrices, respectively. For missing terms of original document-term matrices, zeros are filled in. Besides, the common term vector is also used for the cognitive knowledge map, and zeros are filled in missing terms.

#### *ESOM Clustering*

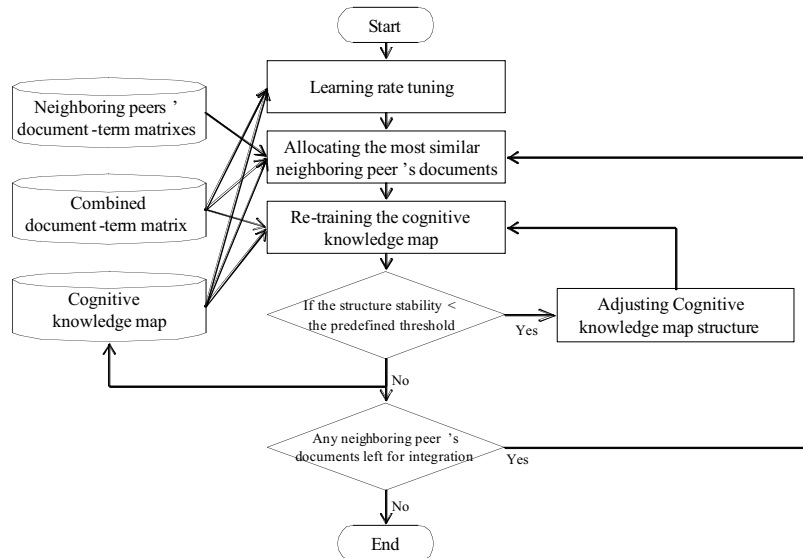
For the purpose of retaining the knowledge structure of the initial knowledge map while integrating the knowledge structure in the cognitive knowledge network, there are three keys of ESOM. Firstly, it provides a parameter, called structure stability, to control the retained degree of the original knowledge structure. Secondly, it uses the smaller learning rate which retains the learning ability without breaking the original knowledge structure as much as possible. Thirdly, it tries to expand the map according to the merged document-term matrix to figure out the more suitable location for the merged document-term matrix. The flowchart of the ESOM algorithm is shown in Figure 3 and described in Section 4.

### **ESOM Clustering algorithm**

#### ***Structure Stability Measurement***

To retain the focal knowledge structure is the major task of ESOM; therefore, we provide the definition of structure stability which is used to measure the retained degree of a trained focal

knowledge structure in this subsection. In this study, we use the SOM-like knowledge map to represent a peer's knowledge structure. Hence, to retain the focal knowledge structure is to retain the focal document clustering result with relative location among clusters in the knowledge map. That is, for each cluster, the clusters upper than it in the original map should be upper than it in the merged map. This condition is also applied to its left side, right side and bottom side clusters. In real world applications, each cluster may contain more than one document; therefore, we use documents instead of clusters to calculate the structure stability and propose a quantification index called structure stability denoted as  $S_i = 1 - (n_r/N)$ , to measure the retained degree of the trained focal knowledge structure accurately, where  $S_i$  is the structure stability of map  $i$ ,  $N$  is the amount of documents represented by the focal map,  $n_r$  is the number of documents which really cause the map's changes. Consequently, structure stability is the proportion of the focal documents which are kept the original relative location among documents. The remaining task for measuring structure stability is how to determine if the relative location change of a document is resulted from itself or other documents. An example is shown in Figure 4. Figure 4(a) shows the document clustering result which only has one document in each cluster and documents are identified by D0 to D8 respectively. Figure 4(b) shows a merged result from Figure 4(a). Although the map size of Figure 4(b) has been changed, the document clustering result and the relative location among clusters are kept. On the contrary, Figure 4(c) is a changed example where the relative location change of D1, D4 and D7 is resulted from D3 and the relative location change of D0, D3 and D6 is resulted from with D4. Thus, D3 and D4 change their corresponding locations with other documents. If we remove D3 and D4, the relative locations of remained documents are right, and we call D3 and D4 *ringleaders*. Therefore, the structure stability of Figure 4(c) is 7/9. In order to figure out ringleaders which actually result in the change of the focal knowledge structure, this study proposes an algorithm as shown in Figure 5.



**Figure 3: The Flowchart of the ESOM Algorithm**

### ***Learning Rate Tuning***

In order to retain the original knowledge structure, ESOM integration process is a fine tuning process. The process of selecting a suitable learning rate is a trial-and-error process. ESOM uses a much smaller number to be the learning rate and the combined document-term matrix to train the cognitive knowledge map. If the structure stability for the trained cognitive

knowledge map is equal to 1, the current learning rate is suitable for being applied in the ESOM. The learning rate tuning can be ascending or descending. In this study, we use the average of the current learning rate and upper limit to be the next learning rate as an ascending strategy, and the average of the current learning rate and lower limit to be the next learning rate as a descending strategy.

(a)

D0	D1	D2
D3	D4	D5
D6	D7	D8

(b)

D0	D1		D2
D3	D4		D5
D6		D7	D8

(c)

D0	D1	D2
D4	D3	D5
D6	D7	D8

**Figure 4: Examples of Structure Stability**

```

1. Before retraining the map, determining and storing the relative locations of all documents.
2. After retraining the map, determining and storing the changed neighbors which are in the
   wrong relative location for all documents.
3. Identifying ringleaders which don't have enough space for it's neighbors,
   Removing these ringleaders from the changed neighboring zone of other documents, and
   Adding these ringleaders to the ringleader list.
4. While (any document with changed neighbors exists) {
   Ringleader  $r$ ;
   Candidate  $c[] \Rightarrow$  documents which have the largest number of changed neighbors;

   If (the number of  $c[]$  is more than one) {
     If (any document which has no legal space for moving) {
        $r \Rightarrow$  selecting a document which has no legal space to move from  $c[]$  randomly;
     } else {
        $r \Rightarrow$  selecting a document from  $c[]$  randomly;
     }
   }

   Removing  $r$  from the changed neighboring zone of other documents;
   Adding  $r$  to the ringleader list;
}

```

**Figure 5: The Pseudo Code of Identifying Ringleaders from a Map**

### ***Allocating the Most Similar Neighboring Peer's Documents***

Conceptually, the neighboring peer's documents which are most similar to current combined documents may have more chances to be directly integrated into the cognitive knowledge map without adjusting the structure of the cognitive knowledge map. Hence, to reduce the complexity of the integration, ESOM has to identify a part of neighboring peer's documents which are most similar to the combined documents in advance and then uses these identified neighboring peer's documents rather than the whole neighboring peer's documents to integrate. The specific process is described as follows. First, ESOM projects all documents of the unmerged neighboring peer's document-term matrix and the combined document-term matrix into the cognitive knowledge map. Next, ESOM calculates the quantization error (QE) for the clusters in which have the unmerged neighboring peer's documents projected. Finally, ESOM selects the cluster which has least QE and uses neighboring peer's documents which are projected in this cluster to be the target for integration.

### ***Re-training the Cognitive Knowledge Map***

In this step, ESOM uses selected neighboring peer's documents and combined documents to train the cognitive knowledge map. The training method is the same as general SOM-like algorithms. Afterward it calculates the structure stability for the trained cognitive knowledge

map. If the structure stability is more than predefined threshold, it means that the integration succeeds. Then, ESOM can further check up if any neighboring peer's documents left for integration. Otherwise, it means that the current cognitive knowledge map has to be adjusted in order to retain focal knowledge structure while integrating neighboring peer's documents.

### ***Adjusting Cognitive Knowledge Map Structure***

Conceptually, the adjustment is to expand the original map size to figure out a more suitable unit, called Unit W, for the selected neighboring peer's documents and therefore these documents will directly be project into Unit W without breaking the original knowledge structure. The clue for figuring out Unit W is the BMU of the selected neighboring peer's document in the original cognitive knowledge map. BMU is the present most suitable unit but still has a gap with Unit W. In order to reduce the gap, ESOM tries to find out the neighboring unit of the BMU which has the most similar weight variation with Unit W. The row/column of units inserted between this neighboring unit and BMU may have more chance to be Unit W. We use the average of the selected neighboring peer's documents to represent Unit W and measure the weight variation between it and neighboring unit of BMU by Eq 1. The unit which has most similar weight variation with D gets the highest score and will be selected. For the new row/column inserted between original units, it directly uses the average of neighboring units and Unit W to initialize. For the new row/column inserted at the boundary of new map, it uses the neighbor in the contrary direction to represent the lacked neighboring unit.

$$Score(i, x) = \sum_{j=1}^n Dif(w_{ij} - w_{xj})$$

$$Dif(w_{ij} - w_{xj}) = \begin{cases} -|w_{ij}|, & \text{if } w_{xj} = 0 \ \& \ w_{ij} \neq 0 \\ 1, & \text{if } w_{xj} = 0 \ \& \ w_{ij} = 0 \\ 0, & \text{if } w_{xj} \neq 0 \ \& \ w_{ij} = 0 \\ \frac{w_{xj}}{w_{ij}}, & \text{if } w_{xj} < w_{ij} \ \& \ w_{xj} > 0 \\ \frac{w_{ij}}{w_{xj}}, & \text{if } w_{ij} < w_{xj} \ \& \ w_{ij} > 0 \end{cases} ,$$

(1)

where  $i$  is the weight of the average of selected neighboring documents and  $x$  is the weight of any one of the neighbors of BMU,  $w_{ij}$  is the weight of  $j^{\text{th}}$  term in  $i$ , and  $w_{xj}$  is the weight of  $j^{\text{th}}$  term in  $x$ .

### **Evaluation of the Visualized Cognitive Knowledge Map Integration System**

We conduct the experiment to evaluate the proposed visualized cognitive knowledge map integration system, *VisCog*, by four criteria.

#### ***Document Set***

The document set used in this experiment contains abstracts of industrial research reports in Chinese gathered from IEK Center, ITRI, Taiwan. There are two reasons for us to choose this document set. First, different from general articles, the abstract of a research report is the essence of original research report; therefore, it is good for extracting knowledge. Secondly, these research reports are published by authors, which we can easily assign abstracts to peers. In order to evaluate the proposed algorithms in details, this study uses 10 peers and 108

abstracts in the evaluation.

### ***Evaluation Criteria***

Three performance criteria, purity, diversity, and specificity are adopted to evaluate the performance of *VisCog* (Agrawal, Bayardo, and Srikant 2000; Wei and Dong 2001; Lin and Hsueh 2006). The definitions of three criteria are described as follows:

$$(1) \text{ Purity} = \sum_{i=1}^m \text{Purity}(i) \times \frac{N_i^U}{N}, \text{ where } \text{Purity}(i) = \frac{n_i^U}{N_i^U}, \text{ } N_i^U \text{ denotes the total number of}$$

documents in the updated cluster  $i$ ,  $n_i^U$  denotes the maximum number of documents that belong to the same category in the updated cluster  $i$ , and  $N$  denotes the total number of documents in all clusters.

$$(2) \text{ Diversity} = \frac{t_U}{T_O} \text{ where } T_O \text{ denotes the number of original category, and } t_U \text{ denotes the number of true categories covered by the updated categories.}$$

$$(3) \text{ Specificity} = \frac{t_U}{T_U}, \text{ where } T_U \text{ denotes the number of updated categories, and } t_U \text{ denotes the number of true categories covered by the updated categories.}$$

Due to the above-mentioned criteria cannot measure the structure change between focal and cognitive knowledge maps, we adopt the structure stability measure described in Subsection 4.1 to measure how much the structure is retained.

### ***Experiment Design***

At first, we take the pre-process as described in subsection 3.1. There are two ways to extract keywords from abstracts: for a peer, it only uses its own documents to extract keywords; for global knowledge map, it uses all documents of all peers to extract keywords. Then, individual peers represent their documents by their document-term matrices and create their individual knowledge maps from their document-term matrices, respectively. Parameters used to create a peer's knowledge map are listed as follows: learning rate = 0.7, tau = 0.6 and training iteration = 10,000. Parameters used to create a global knowledge map are listed as follows: learning rate = 0.7, tau = 0.06 and training iteration = 10,000.

For testing if ESOM can retain whole structure of a focal peer knowledge map, we set the parameter of structure stability as 1 and create the visualized cognitive knowledge map for each peer. Finally, we measure the performance in four criteria for all peers corresponding to the global knowledge map, respectively.

### ***Experimental Results***

In order to evaluate ESOM performance, we generate cognitive knowledge maps for each peer, and calculate four criteria for each peer corresponding to the global knowledge map, respectively. The result of four criteria is shown as Table 2, where we can see the performance of ESOM is not good enough. The possible reason is that the structure stability is set to 1 which means that ESOM will keep the focal structure unchanged, even if the focal structure is not suitable for represent external knowledge distribution. For further examining this assumption, we calculate the Person coefficient of correlation for specificity in cognitive knowledge map generated by ESOM and in the individual knowledge map. The value of correlation coefficient is 0.24, and it clearly shows that if a focal knowledge structure is in better specificity, ESOM can get a better result.

**Table 2: The averages of ESOM Performance on four criteria**

	Purity	Diversity	Specificity	Structure Stability
Peer A	0.85	0.53	0.4	1
Peer B	0.79	0.33	0.63	1
Peer C	0.81	0.33	0.63	1
Peer D	0.93	0.8	0.33	1
Peer E	0.73	0.27	0.33	1
Peer F	0.79	0.27	0.5	1
Peer G	0.72	0.13	0.22	1
Peer H	0.92	0.87	0.25	1
Peer I	0.77	0.2	0.33	1
Peer J	0.97	1	0.13	1
Avg.	0.83	0.47	0.38	1

## Conclusions and Future Research

With the progress of the Internet bandwidth and the personal computer computing capability, Peer-to-Peer (P2P) architecture becomes applicable. Contrast to client-server architecture, P2P architecture more conforms to the autonomous nature of the knowledge workers, and it is more suitable for applying for knowledge management. However, due to the difference between P2P and client-server architectures, knowledge map technologies developed under the client-server architecture may not be directly applied to the P2P architecture. Hence, in order to facilitate knowledge seekers to retrieve other peer's knowledge in P2P networks, this study proposes the visualized cognitive knowledge integration system, called *VisCog*. By using the Egocentric SOM (ESOM) proposed in this study, this system can retain a peer's own knowledge structure while integrating other peer's knowledge structure. In other words, this system can facilitate peers to explore the external knowledge shared by other peers based on its own knowledge structure. Moreover, inheriting SOM-like algorithms, the output of ESOM is a two-dimension topology map which provides the visualization ability for users to trace the mutual relationship between clusters in a plane. Hence, this system can convey more additional information than traditional tree-like models for users to recognize the relationship among peers more effectively.

The main characteristic of ESOM is its process of adjusting the map structure. In a word, it tries to expand the map according to the merged document-term matrix to figure out more suitable locations for the merged document-term matrix. Moreover, it provides peers the ability to determine the retained degree of the original knowledge structure by themselves. A collection of abstracts of industrial research reports from Industrial Economics and Knowledge Center (IEK) is used for evaluate the proposed *VisCog*. A global knowledge map created by GHSOM is used as the reference model to compare with outcomes from ESOM. Three criteria are adopted to measure the purity of updated clusters, the diversity and specificity to denote the recall rate of the original cluster and the precision rate of the updated cluster, respectively. Moreover, in order to measure the knowledge structure remained by ESOM, we propose a new criterion, structure stability, to measure the structure retained in the updated knowledge map.

The results of the evaluation experiments show that ESOM proposed in this study really can retain whole knowledge structure of a focal peer's knowledge map. But the clustering results are very sensitive to a focal peer's knowledge map. In another word, if the knowledge structure of a focal peer is too skewed, the cognitive knowledge map constructed by ESOM

will be skewed, too.

Future research can be continued in following directions: (1) in this version of ESOM, it observes the original knowledge structure inflexibly and lack of a reinforcing learning ability to extract the better knowledge structure to assist clustering. The reinforcement learning is needed for ESOM to adjust individual knowledge structure if needed; (2) in this version of ESOM, it just considers a one-shot integration; for applying in the P2P environment, it may need the ability to merge in an incremental way; (3) parameters for training ESOM may be interrelated; therefore, more comprehensive testing is needed to explore the characteristics of ESOM method.

## References

- Agrawal, R., Bayardo, R., and Srikant, R. "Athena: Mining-based Interactive Management of Text Databases," *Proceedings of the Seventh Conference on Extending Database Technology (EDBT00)*, 2000, pp. 365-379.
- Bonifacio, M., Cuel, R., Mameli, G., and Nori, M. "A Peer-to-Peer Architecture for Distributed Knowledge Management," *Proceedings of the 3rd International Symposium on Multi-Agent Systems, Large Complex Systems, and E-Business (MALCEB' 02)*, 2002.
- Castano, S., Ferrara, A., Montanelli, S., and Zucchelli, D. "HELIOS: a General Framework for Ontology-Based Knowledge Sharing and Evolution in P2P systems," *Proceedings of the 14th International Workshop on Database and Expert Systems Applications*, 2003, pp. 597-603.
- Fritzke, B. "A Growing Neural Gas Network Learns Topologies," *Advances in Neural Information Processing Systems 7*, Tesauero, G., Touretzky, D. S., and Leen, T. K. (ed.), MIT Press, Cambridge MA, 1995, pp. 625-632.
- Lin, F. R., and Hsueh, C. M. "Knowledge Map Creation and Maintenance for Virtual Communities of Practice," *Information Processing and Management* (42:2), 2006, pp. 551-568.
- HUNG, C., and Wermter, S. "A Self-Organising Hybrid Model for Dynamic Text Clustering," *Proceedings of The Twenty-third SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI-2003)*, 2003.
- Kohonen, T. "Self-Organized Formation of Topologically Correct Feature Maps," *Biological Cybernetics* (43), 1982, pp. 59-69.
- Luhn, H. P. "A Statistical Approach to the Mechanized Encoding and Searching of Literary Information," *IBM Journal of Research and Development* (1:4), 1957, pp. 309-317.
- Mangisengi, O., and Essmayr, W. "P2P Knowledge Management: an Investigation of the Technical Architecture and Main Process," *Database and Expert Systems Applications*, 2003, pp. 787-791.
- Rauber, A., Merkl, D., and Dittenbach, M. "The Growing Hierarchical Self-Organizing Maps: Exploratory Analysis of High-Dimensional Data," *IEEE Transactions on Neural Networks* (13:6), 2002, pp. 1331-1341.
- Salton, G., Wong, A., and Yang, C. S. "A Vector Space Model for Automatic Indexing," *Communication of the ACM* (18:11), 1975, pp. 613-20.
- Sparck Jones, K. "A Statistical Interpretation of Term Specificity and its Application in Retrieval," *Journal of Documentation* (28:1), 1972, pp.11-21.
- Wei., C., and Dong, Y. X. "A Mining-based Category Evolution Approach to Managing Online Document Categories," *Proceedings of the 34th Hawaii International Conference on System Sciences*, 2001.